Superpixel-based background removal for accuracy salience person re-identification

Cuong Vo Le*, Quan Nguyen Hong*, Trung Tran Quang*, Nghia Doan Trung†
*School of Electronics and Telecommunications, Hanoi University of Science and Technology,
Hanoi, Vietnam.
Email: cuong.vole@hust.edu.vn
†School of Electrical Engineering, Seoul National University
Email: nghiadt@capp.snu.ac.kr

***Abstract*—This paper presents superpixel-based background removal methods to increase accuracy of global salience person re-identication method. The current algorithm has two problems which limit its accuracy including (1) wrong matching when images of different people have the same background and/or (2) salience on the background of different images of different people is similar. Theoretical maximum accuracy of the global salience method when applying background removal was proved in our previous work by manually remove background close to a human. To achieve an automatic background removal with an accuracy close to the theory, we propose two approaches namely (i) superpixel-GBVS method that combines superpixel and local salience information and (ii) superpixel-pose method that combines superpixel and pose estimation information. The two technologies decide what superpixels belong to human and the others belong to background. It results in a background removal close to edges of the human. Preliminary results show that our results outperform those of the original method without applying any background removal methods and results of the second approach are close to the theory.**

***Keywords*—Person re-identification; Superpixels; Local Salience; Pose Estimation.**

I. Introduction

In recent years, person re-identication has attracted more and more research interests. The aim of this research topic is to identify an individual across non-overlapping camera views. This task faces some challenges such as variation in illumination and pose; and low image resolution. In addition, the non-rigid shape of human body, background clutter and occlusions makes the task non-trivial. Earlier work focused on building relationships between cameras in a network. These methods are known as contextual methods. However, most recent research relies entirely on the analysis of visual descriptors rather than aforementioned contextual information. Based on different points of view, non-contextual methods are broadly classied as passive or active, single-shot or multi-shot, short-period or long-period methods [1]. Some state-of-the-art work focuses on short-period approach which uses only one image for representing the person. Although single-shot methods seem not to be suitable for practical applications, it is considered as a crucial brick in person re-identication since the achievement for this can be directly used to build multi-shot person re-identication.

Recently, Zhao et al. [2] has proposed using saliency information to find the correct matching between an image in the probe and a correspondence in the gallery. This method belongs to the active and single-shot approach and is robust to viewpoint change and pose variation. However, the accuracy of this method is still low because background near the human causes noisy for the re-identification target. We tested Zhao's code [2], in the same condition with the authors', without any background removal method. The average accuracy evaluated on 10 trials at rank 1 on VIPeR dataset is approximately 20%. In parallel, it is also proved that background is one of the main factors which limit matching accuracy in the current method [3]. The similarity of different backgrounds and salience signal on the backgrounds causes wrong matching in many cases. Our previous work [3] proves that the theory maximum Cumulative Matching Characteristic (CMC) accuracy when applying background removal method at rank 1 is 27.18%.
This paper focuses on automatic removing background targeting an accuracy of person re-identification close to the theory. We propose two approaches to remove background namely (1) superpixel and local saliency combination and (2) superpixel and pose estimation combination methods. The two techniques decide what superpixels belong to human and the others belong to background. It results in a background removal close to edges of the human. Through repeating simulation, preliminary results show that our results outperform those of the original method without applying any background removal methods and results of our second approach are close to the theory.

II. Framework

The framework is summarized from Zhao's work with an additional step in the dotted block for automatic background removal, as shown in Fig. 1. The reasons to remove background is that (i) two different people in probe and gallery may have a similar or the same background, as examples in Fig. 2 and/or (ii) two backgrounds of two different people may have a similar saliency, as examples in Fig. 3. One or both of the two problems will result in higher scoring for a matching an image pair. This makes wrong matching of the global salience person re-identification method. There are two approaches for removing the background namely (1) removal by handcraft and

978-1-5090-2743-9/16/$31.00 ©2016 IEEE

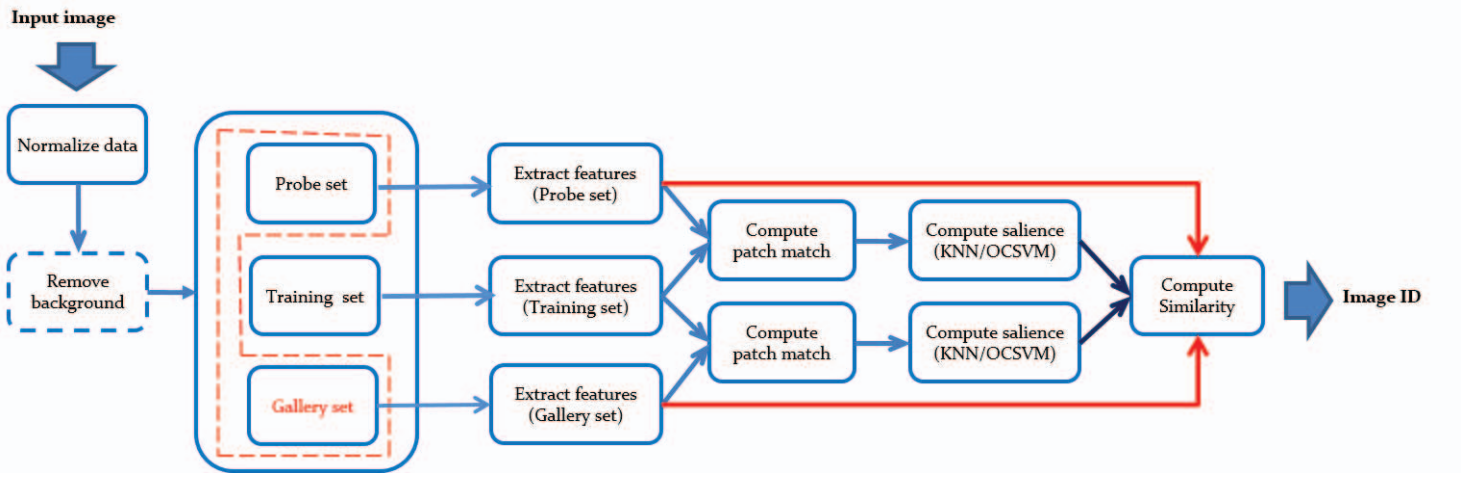

Fig. 1. Diagram of automatic-background-removal based saliency person re-identification

(2) automatic removal. The first way was done by handcraft by using Interactive Segmentation Tool [4] when we proved the theoretical maximum accuracy by background removal method on VIPeR dataset. The method results in background removal very close to human's edges, as examples in Fig. 7(b). However, manual background removal is not applied in a practical system when silhouettes can not be detected. In this paper, we propose two superpixel-based algorithms to automatically remove background close to human's edges. Then, the images after background removal are tested with Zhao's original algorithm to prove the effectiveness of the methods.

A. Theoretically maximum accuracy of background removal

In [3], CMC results for the theoretical maximum accuracy are shown. In this CMC, average increase of all ranks with manual background removal is 20% compared to those in the original saliency person re-identification method, and rank 1 achieves 27.18% on VIPeR dataset.

B. Background removal methods

Several methods are applied in [3]. However, except the manual method, all of them can not remove background close to human's edges. It causes noise to the latter step of person re-identification. We tried to apply an automatic background removal method of optimizing a threshold for a local saliency method named GVBS [5]. However, the method only can achieve accuracy close to the ellipse method but not high enough compared to the theory. To target an accuracy closer to theoretical accuracy when applying background removal method, it is important to remove background as close as possible to human's body edges. Superpixel method [6], [7] is one candidate which can segment a human into many superpixels with all the edges of the human are on the edges of some of the superpixels.

Before applying the superpixel algorithm, a resize step might be required to fit into the current pose algorithm.

Fig. 2. Different people in each pair but similar background from VIPeR

Fig. 3. Similar salient fall into background of images of different people

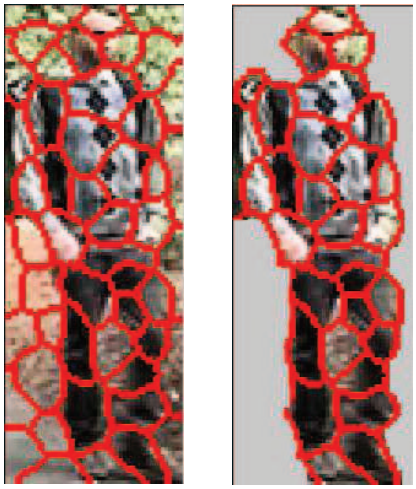

Fig. 4. Image segmentation using Superpixel algorithm: (a) Image segmentation (b) Ideal background removal

Then, 100 superpixels are created on every image and Gibbs (a Markov Chain Monte Carlo algorithm) sampling rate is decided at 5 [6]. Fig. 4 (a) shows an example image after applying the algorithm with some superpixels on the human area and the others on ground area. The next step on the right side is how to keep superpixels that belong to the human and remove the others belonging to the background. Fig. 4 (b) shows an ideal example when it is possible to remove all superpixels on the image's background.

III. Superpixel-based Background Removing Method for accuracy Person Re-identification

The first step is to re-create VIPeR dataset by removing the background on all images in the dataset. The work is based on superpixel method to precisely separate human from the background. The VIPeR dataset after background removal serves as the inputs for the original saliency algorithm for evaluating the accuracy of person re-identification. Two approaches are considered namely (1) superpixel and local salience combination method and (2) superpixel and pose estimation combination method. While salience or pose information decides the raw human area, superpixel will help to detect close to the edges of a human.

A. Background removal by combination of superpixel and saliency

Local saliency is distinctiveness of a patch from its surrounding patches in the image. Graph-Based Visual Saliency (GBVS) is a bottom-up visual saliency model [5]. In this paper, GBVS method is applied to provide a score for every saliency position for each image in VIPeR dataset. It is observed that salience which has high score always appear on the human's area on any image, as shown in Fig. 5 (a). The idea is to keep superpixels which have high saliency score (human) and remove the others (background). However, the range of score is different between images. To make the algorithm effective, every image should have a different saliency threshold. In the VIPeR dataset, after repeating simulation, keeping 50% of superpixels as human and removing 50% of background in an image's bounding box are chosen resulting in a maximum accuracy.

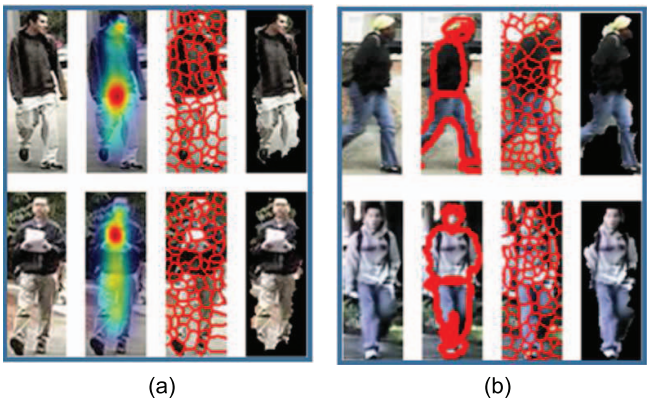

Fig. 5. Results of removing background using (a) Superpixel-GBVS and (b) Superpixel-Pose in the VIPeR dataset

B. Background removal by combination of superpixel and pose

In a similar approach, information after pose estimation [8] is used for keeping superpixels that belong to the human area. This algorithm divides pedestrian images into semantic regions The output is a matrix which has the same size as the original image's size and is divided into regions using a different weight. High weight focuses on the body and the limbs of the people in the photo and low weight will fall into the background. To advance to the results of the manual background removal, we combine the pose estimation method and superpixel to separate close to the edge of human from the background. In practical implementation, some superpixels belong to both pose or non-pose areas needed to be decided if they are human or background. Through repeating simulation for VIPeR dataset, if the area of a superpixel has 65% or greater falling on pose area, it is decided as a part of the human. The chosen parameter achieves maximum accuracy on CMC curve. Fig. 5 (b) is the result of the algorithm on the VIPeR dataset. From left to right, it shows original image, image after pose estimation, applied superpixel image and finally the human separation from background after combining superpixel and pose methods.

C. Evaluation results

1) Simulation conditions: The proposed methods are simulated using VIPeR dataset. We use the same 10 trials as original saliency paper, each has a training set of 218 people and a testing set of 316 people (two images per person). To evaluate the accuracy of person re-identification, CMC curve is used.

2) Simulation result: Fig. 6 shows simulation results of different background removal techniques applied to the global saliency person re-identification. In general, the method by handcraft background removal is on the top as it is the theoretically maximum accuracy. However, our method is better than the two previous methods namely (1) original one and (2) the method only using saliency information for background removal. For the first idea named local-saliency and superpixel combination, until rank 12, it is more effective than the previous one using local salience but not combining with superpixels. For the second approach using pose estimation, the CMC is far from the others and closer to the theory curve. Especially at rank 1, accuracy is 27.03% which nearly reaches to that of the theory method of 27,18%. The detailed comparison in different ranks is shown in Table 1. Fig. 7 shows an example of the effectiveness of our method. For the original method, the right matching person is at rank 8 in the gallery set but it changes to rank 1 for both of our two proposed methods.

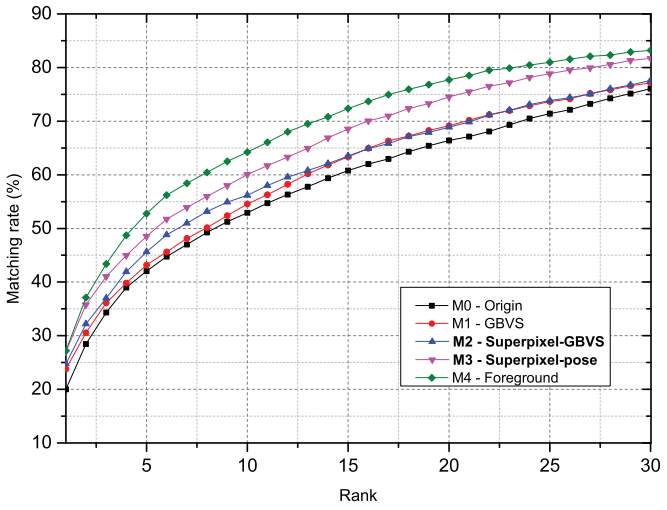

Fig. 6. Comparison of different person re-identification methods.

TABLE I
MATCHING RATE OF DIFFERENT METHODS (%)

Rank	1	5	10	15	20	25	30
M0 [2]	20.00	42.06	52.91	60.69	66.39	71.39	76.08
M1 [3]	23.80	43.20	54.56	63.35	69.15	73.67	77.12
M2	24.75	45.63	56.17	63.51	68.89	73.82	77.59
M3	27.03	48.54	60.06	68.54	74.53	78.86	81.71
M4 [3]	27.18	52.78	64.24	72.34	77.69	80.89	83.20

IV. CONCLUSION

In this paper, we propose two approaches namely superpixel-GBVS and superpixel-pose to remove background from all images on VIPeR dataset. Then, the images are tested with global saliency method proving the effectiveness

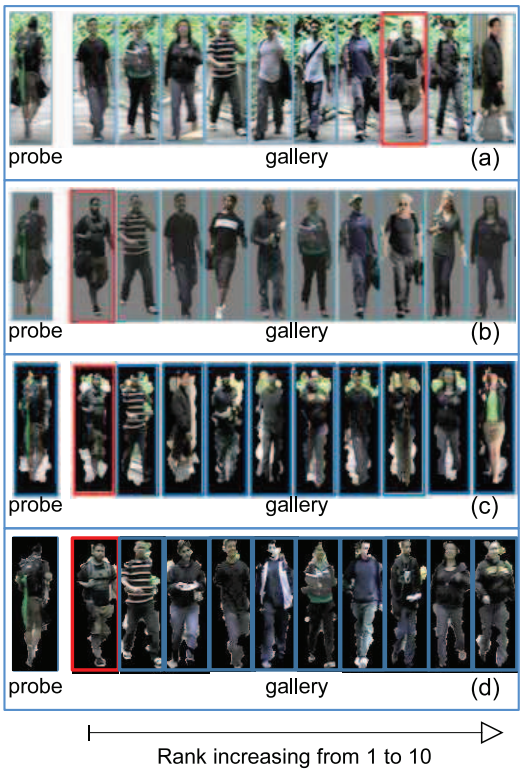

Fig. 7. Example of 10 first rank results of the probe image for (a) original method; (b) interactive segmentation; (c) superpixels local salience; (d) superpixels pose

of the automatic background removal method. Parameters of the system are chosen through repeating simulation. It shows that the two proposed methods outperform some previous published methods. Our second approach using combination of superpixel and pose estimation has results close to the theory CMC curve. Especially at rank 1, the accuracy of out second method reaches that of the theory of 27.03%.

REFERENCES

[1] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *in Image and Vision Computing*, vol. 32, pp. 270-286, no. 4, 2014

[2] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identication," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3586-3593, 2013.

[3] T. B. Nguyen, V. P. Pham, T. Le, and C. V. Le, "Background removal for improving saliency-based person re-identification," to be appeared in PR4MCA, 2016.

[4] *http://kspace.cdvp.dcu.ie/public/interactive-segmentation*

[5] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *in Advances in neural information processing systems*, pp. 545-552, 2006.

[6] Greg Mori, "Guiding Model Search Using Segmentation," *IEEE International Conference on Computer Vision*, 2005.

[7] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering Human Body Configurations: Combining Segmentation and Recognition," *IEEE Computer Vision and Pattern Recognition*, 2004.

[8] P. Luo, X. Wang, and X. Tang, "Pedestrian Parsing via Deep Decompositional Neural Network," *in Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013.